Translational control of gene expression via interacting feedback loops

Wang, Liang; Romano, M. Carmen; Davidson, Fordyce A.

Published in:
Physical Review E: Statistical, nonlinear, and soft matter physics

DOI:
10.1103/PhysRevE.100.050402

Publication date:
2019

Document Version
Peer reviewed version

Translational control of gene expression via interacting feedback loops

Liang Wang
Division of Mathematics, University of Dundee, DD1 4HN, UK

M. Carmen Romano
SUPA, Institute for Complex Systems and Mathematical Biology, Department of Physics, Aberdeen AB24 3UE, UK and Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen AB24 3FX, UK

Fordyce A. Davidson
Division of Mathematics, University of Dundee, DD1 4HN
*Corresponding Author**
(Dated: March 10, 2020)

Translation is a key step in the synthesis of proteins. Accordingly, cells have evolved an intricate array of control mechanisms to regulate this process. By constructing a multi-component mathematical framework we uncover how translation may be controlled via interacting feedback loops. Our results reveal that this interplay gives rise to a remarkable range of protein synthesis dynamics, including oscillations, step-change and bistability. This suggests that cells may have recourse to a much richer set of control mechanisms than was previously understood.

Keywords: translation, feedback, TASEP, oscillation, bistability

Control of gene expression refers to the processes by which the production of proteins is regulated by the cell. This is at the heart of the functioning of all living organisms and it allows cells to adapt to their environment. Control of gene expression can occur at multiple levels [1, 2]. In this paper we focus on translational control.

Translation is the process by which a protein is made from a messenger RNA (mRNA) molecule. An mRNA consists of a sequence of codons, each coding for a certain amino acid. Translation is performed by molecular machines called ribosomes, which bind to the beginning of the mRNA (5' UTR region), scan it for the start codon and hop from one codon to the next, thereby producing the chain of amino acids which form the protein. When the ribosome reaches the stop codon, the protein is complete, is released into the cytoplasm and the ribosome binds off the mRNA.

Recent years have witnessed an explosion of information about how translational mechanisms regulate protein levels [3]. Prominent examples include translational control during cell stress [4], switching in the mechanism responsible for translation initiation during the cell cycle [5], and translational repression by microRNAs [6–9]. One of the main advantages of translational control is that allows for a rapid cell response [10].

In this paper, we focus on one important case of translational control that has remained unexplored within this research framework, namely the interplay between positive and negative regulatory mechanisms. Proteins that bind to the 5'UTR region of mRNAs and hinder translation play an important role in regulation of gene expression [18–20]. Examples include proteins involved in neural differentiation and plasticity [21], cognitive problems [22], cell cycle and differentiation of mammalian cells [23, 24] and DNA repair [25]. On the other hand, virtually all mRNAs are subject to positive feedback via ribosome recycling due to their pseudo-circular structure [26, 27]. A particularly pertinent example is the Poly(A) binding protein (PABP) [28–32]. PABP binds to the 3' end (poly(A) tail) of all mRNAs and strongly affects both translational efficiency and stability of all transcripts. Moreover, PABPs interact with the initiation factors bound to the cap of the mRNA, facilitating the circularisation of the mRNA and consequent recycling of ribosomes [26, 33]. Finally, PABPs protect the mRNA from degradation [34]. Therefore, PABPs exert control on protein synthesis via a two-fold mechanism, and as consequence, regulation of PABP availability provides a global mode of translational regulation.

Here we show that the interplay between negative translational feedback and ribosome recycling gives rise to a novel range of dynamical behaviour in protein synthesis, including oscillations, step-change and bistability. The existence of at least one negative and one positive feedback loop has been proved to be a necessary condition for oscillatory behaviour and multistability, respectively [11–15], but examples of this dynamical behaviour have been studied mainly in transcriptional regulation [16] or coupling between transcription and translation [17]. To our knowledge, the generation of oscillatory and bistable behaviour by purely translational regulatory mechanisms has not been studied before.

Modelling framework.–Our mathematical framework is a multi-component model that accounts for transla-

* f.a.davidson@dundee.ac.uk

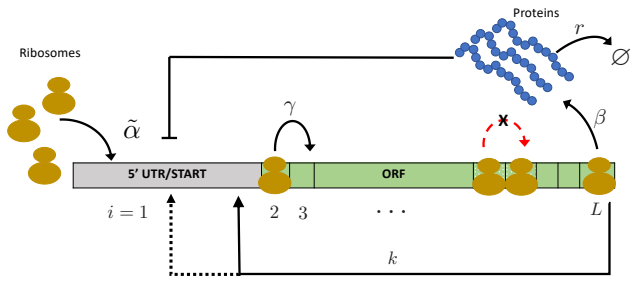

FIG. 1. Schematic of translation model with ribosome recycling and auto-negative feedback. Competitive recycling (dashed line), non-competitive recycling (solid line). Each site can be occupied by no more than one particle, so that at any time t the state at site i given by $S_i(t)$ is either $S_i(t) = 0$ or $S_i(t) = 1$, with $i = 1, ..., L$, where L is the lattice length. Particles bind to the first site of the lattice at rate $\tilde{\alpha}$, then hop from one site to the next at rate γ (usually rescaled to one and done so here) and finally leave the lattice from the last site at rate β, marking the point where the associated protein synthesis is completed. See text for further details.

tion, protein complex formation and binding of protein complexes and ribosomes at the 5'UTR (see Fig. 1). At its core is a stochastic model of one-dimensional transport extensively studied in non-equilibrium statistical physics - the Totally Asymmetric Simple Exclusion Process (TASEP) [35]. Here we augment this standard model with processes that explicitly account for protein synthesis and associated positive and negative feedbacks. This allows us to consider both steady state and dynamic behaviours. Our model predicts the average number of particles per site ρ and the consequent production of proteins. Ribosomes are assumed to attach at a rate $\tilde{\alpha}$ and move along the mRNA at a rate γ. On reaching the final site, ribosomes leave with rate β and are recycled at rate k. At this stage a protein is produced. The protein number N is controlled by a balance of synthesis and degradation with rate constant r. Critical to the model is that both proteins and ribosomes feed back to affect the loading rate $\tilde{\alpha}$ (see Fig. 1). We consider lattices representing realistic mRNA lengths of 500 codons and simulation times that are of the order of the half-life of typical prokaryotic mRNAs (see Supp. Mat.).

Translation. Ribosomes are represented by particles that hop stochastically along the sites of a one dimensional lattice corresponding to the codons of the mRNA [36]. Note that in the standard case (setting recycling and feedback to zero here) steady state traffic on the lattice can be classified into three main phases: *low density* (LD) ($\alpha < \beta, \alpha < 1/2$), *high density* (HD) ($\beta < \alpha, \beta < 1/2$) and the *maximal current* (MC) ($\alpha, \beta \geqslant 1/2$). Each of these phases has an associated average density ρ and current J (average flux of particles), which in the limit of an infinitely long lattice are given by: $\rho_{LD} = \alpha, \rho_{HD} = 1 - \beta, \rho_{MC} = 1/2$ and $J_p = \rho_p(1 - \rho_p)$, $p \in \{LD, HD, MC\}$ [37, 38]. Corresponding expressions for the modified TASEP considered here will be presented below.

Translational negative feedback. A protein can bind (often in multimeric form) to its own mRNA, thereby blocking the loading of ribosomes. Since protein binding/unbinding to the mRNA is generally much faster than ribosome loading [39], the probability of the start codon being free for ribosome loading can be described by a Hill-function $f(N) = 1/(1 + (N/\theta)^n)$, where N is the protein copy number, θ measures the protein level that induces half maximal ribosome binding rate and n measures cooperativity of the protein multimer (see Supp. Mat.). Thus, the intrinsic initiation rate is modified from α (the standard constant rate) to $\alpha f(N)$.

Translational positive feedback. The two ends of the mRNA can interact leading to a pseudo-circular structure [40], which together with the recycling complex Rli1p [41] promote terminating ribosomes to start a new round of translation on the same mRNA [42]. Following [43], a ribosome on site $i = L$ is assumed to either detach at rate β and enter the reservoir of free ribosomes or move directly onto site $i = 1$ at a *recycling rate* k (if $S_1(t) = 0$) to re-initiate the translation process.

Protein degradation. Once synthesised, proteins enter the intra-cellular pool, where they are subjected to degradation. This can be a complex process [44, 45]. However, as a detailed description is not critical to the work presented, we adopt the widely used approach of modelling removal as a Poisson process with resultant removal rate rN.

Finally, note that if the processes detail above are in steady state, then $J = rN$ and hence $N \equiv N^* = J/r$ and hence we can write $f = 1/(1 + (4I\,J)^n)$, where we have introduced the reciprocal factor $I := 1/(4\theta r)$ that measures *feedback intensity* (the factor of 4 is for algebraic convenience).

Models for interacting feedback loops. Experimental results suggest that recycled ribosomes are channelled downstream of the normal *de novo* initiation site and thus may evade the blocking effect of the protein complex [46]. This is the case discussed here and referred to as *non-competitive recycling*. However, the relative position of the protein complex binding site and the recycled ribosome initiation site is not clear. Hence, an alternative is that both recycled ribosomes and *de novo* initiation are blocked by the protein complex, (*competitive recycling*) (see Fig. 1). We comment on this alternative case below.

In the *non-competitive recycling* case a system of ordinary differential equations that determine the dynamics

of the average occupancies ρ_i of the lattice sites is:

$$\begin{aligned}\frac{d\rho_1}{dt} &= \underbrace{\alpha f(N)(1-\rho_1)}_{de\ novo} + \underbrace{k\rho_L(1-\rho_1)}_{\text{recycled}} - \rho_1(1-\rho_2),\\ \frac{d\rho_i}{dt} &= \rho_{i-1}(1-\rho_i) - \rho_i(1-\rho_{i+1}), \quad i = 2,\dots,L-1,\\ \frac{d\rho_L}{dt} &= \rho_{L-1}(1-\rho_L) - \beta\rho_L - \underbrace{k(1-\rho_1)\rho_L}_{\text{recycled}}.\end{aligned} \tag{1}$$

By direct comparison with the corresponding system for the standard TASEP, *effective* entry and exit rates can be defined as follows

$$\alpha_{eff} := \alpha f(N) + k\rho_L, \quad \beta_{eff} := \beta + k(1-\rho_1). \tag{2}$$

Effective rates for the *competitive recycling* case can be defined equivalently (see Supp. Mat.). However, the behaviour of these two cases is similar in almost all parameter regimes and we present only the results for the non-competitive case highlighting where any differences in the two cases arise.

Steady-state analysis of protein production and ribosome density.– Setting the right hand side of System 1 to zero and applying a mean-field approach leads to conditions that partition steady states of the system into phases aligned with those for the standard case (LD, HD and MC). These conditions are defined by direct substitution of α and β with α_{eff} and β_{eff}, respectively, in the standard conditions stated above. In turn and after some lengthy analysis, expressions for these phases and their boundaries in terms of the system parameter are as follows - note these definitions collapse to the standard cases on setting $k = I = 0$ (see Supp. Mat.).

Maximal Current Phase

$$\frac{\alpha}{2\alpha + k\,(1+I^n)} \leqslant \begin{cases} \beta & \leqslant \dfrac{\alpha k}{1+I^n-2\alpha} \quad \alpha < \dfrac{1}{2}\,(1+I^n)\,, \\ \beta & \text{otherwise.} \end{cases} \tag{3}$$

Within this region,

$$\alpha_{eff} = \frac{\alpha}{1+I^n} + \frac{k}{2\left(\beta + \sqrt{\beta\left(\beta + \dfrac{k}{\alpha}\,(1+I^n)\right)}\right)} \quad \text{and}$$

$$\beta_{eff} = \frac{1}{2}\left(\beta + \sqrt{\beta\left(\beta + \frac{k}{\alpha}\,(1+I^n)\right)}\right).$$

Low Density Phase

$$\alpha < \begin{cases} \beta\left(1 + \left(\dfrac{4I(\beta+k)(1-\beta)}{(1+k)^2}\right)^n\right) & 2\beta + k < 1, \\ \dfrac{\beta}{2\beta+k}\,(1+I^n) & \text{otherwise.} \end{cases} \tag{4}$$

Within this region, there exists a unique, positive expression for α_{eff} that yields unique, positive expressions for β_{eff} and N^*.

High Density Phase

$$\alpha > \begin{cases} \beta\left(1 + \left(\dfrac{4I(\beta+k)(1-\beta)}{(1+k)^2}\right)^n\right) & 2\beta + k < 1, \\ \dfrac{k\beta}{1-2\beta}\,(1+I^n) & \text{otherwise.} \end{cases} \tag{5}$$

Within this region, there exist either one, two or three positive solutions, β_{eff}, that yield positive expressions for α_{eff} and N^*.

For $k, I > 0$ and for a large range of parameter space, long run Monte-Carlo simulations of the modified TASEP reveal steady state phases that are well-characterised by Eqs. (3)-(5) (see Figs. S1 and S2). However, a deeper analysis reveals that the complex interplay of positive and negative feedback can generate distinct and novel *dynamical* behaviour.

Negative feedback and ribosome recycling induce oscillations in cellular protein level.– Monte-Carlo simulations reveal periodic oscillations in the number of proteins $N(t)$ within the initiation limited regime (LD phase). The stochastic nature of the individual simulations makes it difficult to systematically differentiate periodic oscillations from random fluctuations by visual inspection. However, a power spectrum analysis provides a clear demarcation: a tight, single-peaked spectrum is associated with periodic oscillations (Figs. 2A,C) whereas a broad band response is obtained in the case of stochastic fluctuations (Figs. 2B,D).

The time needed for a ribosome to transit the mRNA induces a delay between initiation and completion of protein synthesis. This generates a delay in the action of the negative feedback - a mechanism known to generate oscillatory behaviour [47]. A simplified model for the protein copy number $N(t)$ in the LD regime is:

$$\begin{aligned}\frac{dN(t)}{dt} &= J(t) - rN(t)\\ &= \alpha_{eff}(t-T)\big(1-\alpha_{eff}(t-T)\big) - rN(t),\end{aligned} \tag{6}$$

where T denotes the translational delay time [48]. Appealing to Eqs. 2 and setting $\rho_L = J(t)/\beta_{eff}$, it follows that

$$\alpha_{eff}(t) = \frac{\alpha(\beta+k)}{\alpha k + \beta\Big(1 + \big(4IrN(t)\big)^n\Big)}. \tag{7}$$

Substituting Eq. 7 into Eq. 6 results in a delay differential equation for N. The translational delay can be estimated as $T = L/(1-\rho^*)$, where $\rho^* := \alpha_{eff}(N^*)$. This simplified model reproduces the amplitude and period of the stochastic simulations (cf. Fig. 2 A and Fig. S3D). It can be shown that on increasing α the steady state of (6) can be driven unstable via a Hopf bifurcation with

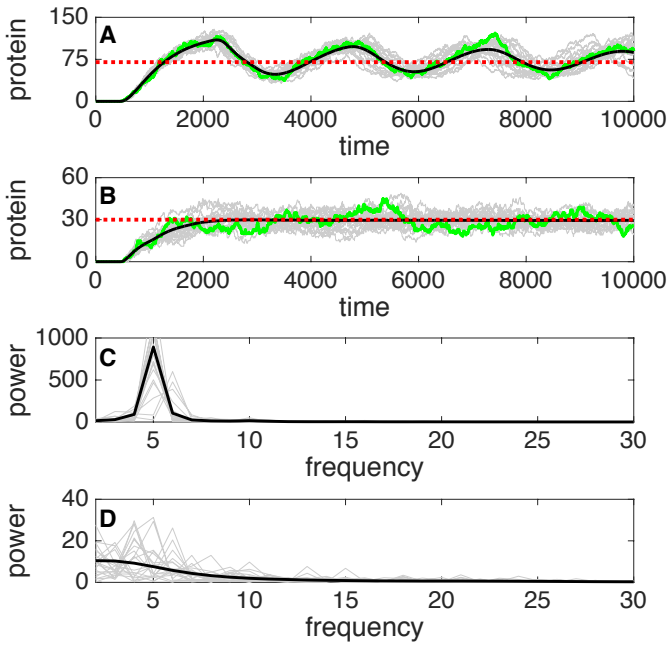

FIG. 2. Simulation and power spectrum of protein level in low density phase. (A,B) Protein number as a function of time. Averages over 5,000 realizations of stochastic simulations (black lines), 20 example realizations (grey lines), single example trajectory (green line). Red dashed lines are mean field N^* as computed using the steady state theory. Typical simulation (A) and the corresponding power spectrum (C) from the region where oscillation are predicted to exist ($\alpha = 0.8$; note $\alpha_{eff} = 0.17 < 0.5$ therefore corresponding to the system being in LD phase). Typical simulation (B) and the corresponding power spectrum (D) from the region where oscillation are not predicted to occur ($\alpha = 0.05$). In all cases $\beta = 0.5,\ k = 0.2,\ I = 5/2,\ r = 0.002,\ n = 5,\ L = 500$. 1 time unit = 1/22 secs.

the Hopf locus given by

$$B\cos(\sqrt{B^2 - r^2}\,T) + r = 0, \tag{8}$$

where B is a function of the system parameters (see Supp. Mat.). This locus forms a curve in the $\alpha - \beta$-plane (see Fig. 3). After some algebra, it can be shown that necessary conditions for the existence of the Hopf locus are $n > 1$ and $I > F(\alpha, \beta, k, n)$ for some positive function F (see Supp. Mat.). The condition $n > 1$ indicates that cooperativity in protein binding is necessary for the onset of oscillations [49]. The second condition indicates that the onset of oscillations occurs when the feedback intensity is sufficiently strong.

As one would intuitively expect, increasing feedback intensity induces onset of oscillations at lower values of the intrinsic loading rate α (see Fig. 3A), Interestingly, the Hopf locus also shifts left on increase the recycling rate (see Fig. 3B). Hence, counterintuitively, ribosome recycling - a positive feedback mechanism - also *enhances* the onset of oscillations [50].

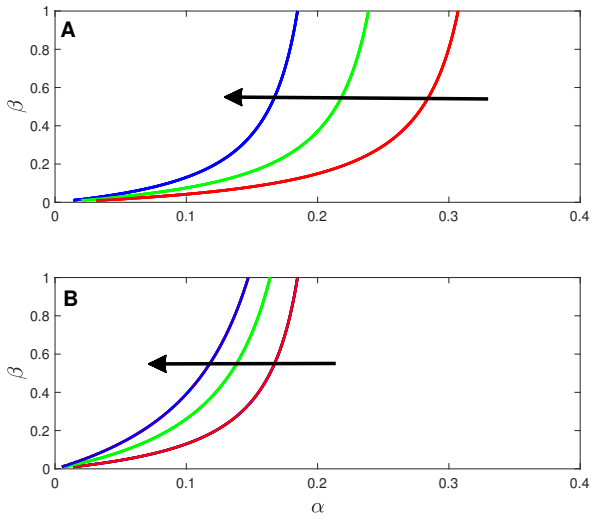

FIG. 3. The Hopf locus and the effect of varying recycling and feedback. Hopf bifurcation locus in the $\alpha - \beta$ plane computed using Eq. (8). The effect of varying feedback intensity (A) and recycling (B) is shown. Black arrows indicate direction of increasing effect. (A) $I = 25/14$ (red), $I = 25/12$ (green), $I = 25/10$ (blue) with $k = 0.2$ (B) $k = 0.2$ (red), $k = 0.4$ (green), $k = 0.6$ (blue) with $I = 25/10$. $L = 500$ sites, $r = 0.002$, $n = 5$.

Interplay between recycling and negative feedback induces bistability in protein production. –In the MC and LD phases, the current J is uniquely defined for any given parameter set. On the contrary, in the HD phase J can be multiply defined: after some algebra it can be shown that β_{eff} is the solution to the following $2n + 1$ degree equation (see Supp. Mat.):

$$4^n k I^n \beta \beta_{eff}^n \left(1 - \beta_{eff}\right)^{n+1} - (k\beta + \alpha)\beta_{eff} + \beta(\alpha + k) = 0. \tag{9}$$

For $k = 0$, Eq. (9) has the unique solution $\beta_{eff} = \beta$. For $I = 0$, β_{eff} is uniquely defined by $\beta_{eff} = \beta(\alpha + k)/(k\beta + \alpha)$ [43]. However, when both $k,\ I > 0$, Eq. (9) can have three admissible solutions, depending on the value of α. Thus, for suitably chosen parameters, there exists an interval of values of α for which three steady state values of $N^* = J/r$ co-exist. Figure 4A shows N^* as function of α, so that on increasing α from zero to one, the model transits from the LD to the HD phase (at $\alpha \approx 0.5$).

With values of α selected from the co-existence interval shown in Fig. 4A, simulations reveal that the number of proteins remains centred on high or low states for a time scale orders of magnitude larger than stochastic fluctuations (4B). Rapid switching events between the favoured state are accompanied by a brief hiatus at an intermediate state. The mean locations of these favoured and intermediate states are well-approximated by the an-

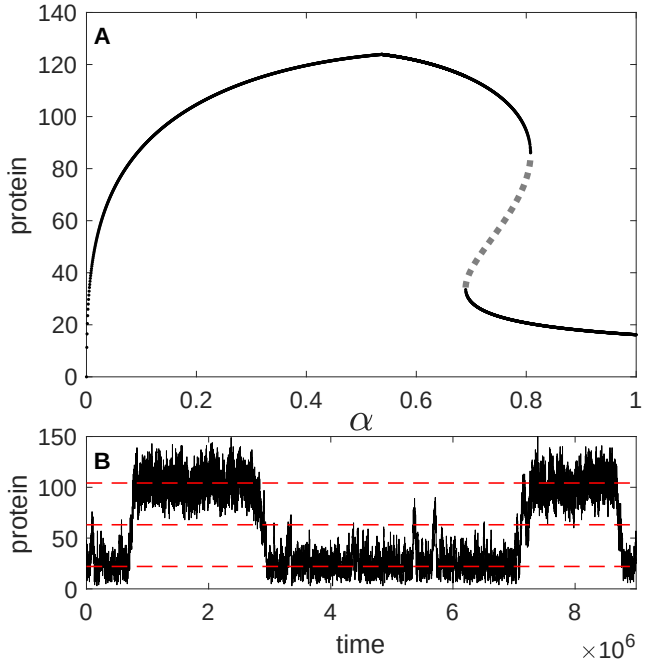

FIG. 4. (A) Protein number N^* as a function of the initiation rate α as predicted by the steady state theory. N^* is first monotonically increasing (LD phase) then decreasing (HD phase) [43]. N^* then passes through two fold bifurcations, leading first into, and then out of the interval of co-existent states - with the upper and lower branches (solid black lines) separated by an intermediate state indicated (dashed grey line). (B) Monte-Carlo simulation of protein number a function of time (solid black line) with the mean field solutions N^* (red dashed lines) from A ($\alpha = 0.77$). In both cases $\beta = 0.015$, $k = 0.8$, $I = 6$, $r = 0.002$, $n = 2$, $L = 500$. 1 time unit $= 1/22$ secs.

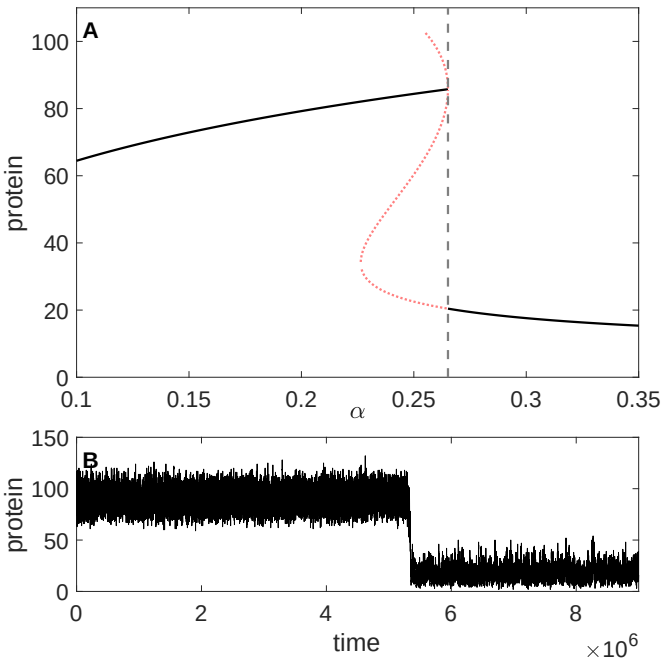

FIG. 5. Step-change in steady state protein levels. (A) Signal-response curve for protein number N^* as a function of the initiation rate α (black curves). Inadmissible solutions for (9) red dotted curve. The LD-HD boundary vertical grey dashed line. (B) Time series of the number of proteins: $\alpha = 0.28$ then switched to $\alpha = 0.3$ at time $t = 4.5 \times 10^6$. In each case $\beta = 0.015$, $k = 0.263$, $I = 6$, $r = 0.002$, $n = 2$, $L = 500$. 1 time unit $= 1/22$ secs.

alytic expressions for the steady states obtained from Eq. 9 (Fig. 4A). Frequency histograms reveal the effect of varying α across the bistable region and together with dwell-time histograms, indicate this to be a memory-less stochastic switching process (see Figs. S5 and S6)

Fixing k (resp. I) and increasing I (resp. k) increases the interval of values of α for which the fold exists (fold width - see Fig. S7). Interestingly, the location of the fold is also an increasing function of I and k. Indeed, somewhat counter-intuitively, for a fixed value of α, increasing the intensity of the negative feedback I can force the system from a low to high N^* state. To understand this recall that within the HD regime, any change of parameters leading to a decrease in the ribosome density leads to an increase in the ribosomal current: increasing I, decreases α_{eff}, thereby decreasing the density of ribosomes on the mRNA. Finally, we note that bimodality in protein production rate is a result of a careful balance between the negative and positive feedback loops and tuning one or the other can drive the system both into and out of a bimodal response (see Fig. S7). Indeed, in the case of competitive recycling no bistability occurs. Following similar calculations, it can be shown that the equation corresponding to Eq. 9 has a unique solution. It appears that feedback dominates and locks the system into a unique steady state.

Feedback interplay can induce step-changes in protein production–If k is fixed to ensure bistability, a critical value of I exists at which the right boundary of the bistable region coincides precisely with the LD-HD boundary. In this case, as the initiation rate α passes through the LD-HD boundary, a discontinuity in the number of proteins occurs (Fig. 5A). [Qualitatively similar behaviour is obtained on keeping I fixed and varying k.] This step-change in the number of proteins can be large, suggesting that small changes in the ribosome initiation rate α can result in a significant shift in protein levels. Simulations confirm this theoretical prediction. On increasing α dynamically during a simulation, a step change (around 75% reduction) was clearly induced on transept of the LD/HD critical value (Fig. 5B). This cliff-edge response is another unique feature of resulting from the interplay between feedback and recycling.

Conclusions.–Our model shows the rich dynamical behaviour caused by the interplay of negative and

positive feedback loops in translation, and it provides a general mathematical framework to analyse other kinds of gene expression regulation, such as regulation exerted by miRNAs [51]. It suggests that this interplay could provide cells with a versatile mechanism to adapt their protein levels according to the environment. The centrally important protein PABP is subject to ribosome recycling and is known to exhibit translational negative feedback. Interestingly, PABP has also been implicated in circadian oscillations [52]. Finally, disturbances of poly(A) tail length have been linked to a number of physiological and pathological processes. Therefore, a better understanding of the interplay of ribosome recycling and translational negative feedback has far reaching consequences.

Acknowledgements. LW was supported by the Northern Research Partnership

[1] J. Hausser, L. Keren, and U. Alon, Nature Communications **10** (2019).

[2] U. Alon, *An introduction to systems biology: design principles of biological circuits* (Boca Raton, FL, Chapman & Hall/CRC, 2007).

[3] J. Hershey, N. Sonenberg, and M. Mathews, Cold Spring Harb Perspect Biol. **4**, a011528 (2012).

[4] K. Spriggs, M. Bushell, and A. Willis, Mol. Cell **40**, 228 (2014).

[5] S. Pyronnet, J. Dostie, and N. Sonenberg, Genes Dev. **15**, 2083 (2001).

[6] M. Selbach, B. Schwanhusser, N. Thierfelder, Z. Fang, R. Khanin, and N. Rajewsky, Nature **455**, 58 EP (2008).

[7] A. Zinovyev, N. Morozova, A. N. Gorban, and A. Harel-Belan, "Mathematical modeling of microrna–mediated mechanisms of translation repression," in *MicroRNA Cancer Regulation: Advanced Concepts, Bioinformatics and Systems Biology Tools*, edited by U. Schmitz, O. Wolkenhauer, and J. Vera (Springer Netherlands, Dordrecht, 2013) pp. 189–224.

[8] S. Nikolov, J. V. Gonzalez, M. Nenov, and O. Wolkenhauer, Biotechnology & Biotechnological Equipment **26**, 3315 (2012).

[9] A. Zinovyev, N. Morozova, N. Nonne, E. Barillot, A. Harel-Bellan, and A. N. Gorban, BMC Systems Biology **4** (2010), 10.1186/1752-0509-4-13.

[10] J. W. Hershey, N. Sonenberg, and M. B. Mathews, Cold Spring Harbor Perspectives in Biology (2018), 10.1101/cshperspect.a032607.

[11] J. L. Gouz, Journal of Biological Systems **06(1)**, 11 (1998).

[12] E.-H. Snoussi, Journal of Biological Systems **06(1)**, 3 (1998).

[13] P. François and V. Hakim, Proceedings of the National Academy of Sciences **101**, 580 (2004), https://www.pnas.org/content/101/2/580.full.pdf.

[14] L. Potvin-Trottier, N. Lord, G. Vinnicombe, and J. Paulsson, Nature **538**, 514?517 (2016).

[15] E. Ullner, A. Koseska, J. Kurths, E. Volkov, H. Kantz, and J. García-Ojalvo, Phys. Rev. E **78**, 031904 (2008).

[16] R. Hermsen, B. Ursem, and P. R. ten Wolde, PLOS Computational Biology **6**, 1 (2010).

[17] Y. Zarai and T. Tuller, PLOS Computational Biology **14**, 1 (2018).

[18] K. Abdelmohsen, *Modulation of Gene Expression by RNA Binding Proteins: mRNA Stability and Translation Fundamental Principles of Optical Lithography* (InTech Open, 2012).

[19] O. Levi and Y. Arava, PLOS Biology **17**, 1 (2019).

[20] R. Betney, J. de Silva, and I. Stansfield, RNA-A **16**, 655 (2010).

[21] H. M. N and L. H, Cell Mol Life Sci **65(20)**, 3168 (2008).

[22] G. J. Bassell and S. T. Warren, Neuron **60**, 201 (2008).

[23] C. K. Damgaard and J. Lykke-Andersen, Genes & Development **25**, 2057 (2011).

[24] V. Dormoy-Raclet, J. Markovits, A. Jacquemin-Sablon, and H. Jacquemin-Sablon, RNA Biology **2**, 112 (2005).

[25] T. Fukuda, M. Ashizuka, T. Nakamura, K. Shibahara, K. Maeda, H. Izumi, K. Kohno, M. Kuwano, and T. Uchiumi, Nucleic Acids Research **32**, 611 (2004).

[26] S. Wells, P. Hillner, R. Vale, and A. Sachs, Mol. Cell **2**, 135 (1998).

[27] Z. A. Afonina, A. G. Myasnikov, V. A. Shirokov, B. P. Klaholz, and A. S. Spirin, Nucleic Acids Research **42**, 9461 (2014), http://oup.prod.sis.lan/nar/article-pdf/42/14/9461/14123029/gku599.pdf.

[28] E. Hornstein, H. Harel, G. Levy, and O. Meyuhas, FEBS Letters **457**, 209 (1999).

[29] O. Neto, N. Standart, and C. Desa, Nucleic Acids Research **23**, 2198 (1995).

[30] J. Bag, J. Biol. Chem. **276**, 47352 (2001).

[31] G. Patel, S. Ma, and J. Bag, Nucleic Acids Res. **33(22)**, 7074?7089 (2005).

[32] B. Gorgoni and N. Gray, BRIEFINGS IN FUNCTIONAL GENOMICS AND PROTEOMICS **3**, 125?141 (2004).

[33] A. Kahvejian, Y. V. Svitkin, R. Sukarieh, M. N. M'Boutchou, and N. Sonenberg, Genes & Development **19**, 104 (2005).

[34] D. Cao and R. Parker, RNA **7**, 1192 (2001).

[35] B. Derrida, E. Domany, and D. Mukamel, J. Stat. Phys. **69**, 667 (1992).

[36] C. MacDonald, J. Gibbs, and A. Pipkin, Biopolymers **6**, 1 (1968).

[37] B. Derrida, M. Evans, C. Hakim, and V. Pasquier, J. Phys. A: Math. Gen. **26**, 1493 (1993).

[38] G. Schütz and E. Domany, J. Stat. Phys. **72**, 277 (1993).

[39] L. Chen, R. Wang, T. Kobayashi, and K. Aihara, Phys. Rev. E **70**, 011909 (2004).

[40] D. Barthelme, S. Dinkelaker, S.-V. Albers, P. Londei, U. Ermler, and R. Tampe, Proc. Natl Acad. Sci. **108**, 3228 (2011).

[41] C. Shoemaker and R. Green, Proc Natl Acad. Sci. **108**, 1392 (2011).

[42] N. Amrani, S. Ghosh, D. Mangus, and A. Jacobson, Nature **435**, 1276 (2008).

[43] E. Marshall, I. Stansfield, and M. Romano, J. R. Soc. Interface **11**, 104 (2014).

[44] D. A. Rothenberg, M. J. Taliaferro, S. M. Huber, T. J. Begley, P. C. Dedon, and F. M. White, iScience **9**, 367

(2018).
[45] R. Golan-Lavi, C. Giacomelli, G. Fuks, A. Zeisel, J. Sonntag, S. Sinha, W. Kstler, S. Wiemann, U. Korf, Y. Yarden, and E. Domany, Cell Reports **18**, 3129 (2017).
[46] L. Rajkowitsch, C. Vilela, K. Berthelot, C. Ramirez, and J. McCarthy, J. Mol. Biol. **335**, 71 (2004).
[47] S. Pigolotti, S. Krishna, and M. Jensen, Proc Natl Acad. Sci. **104**, 6533 (2007).
[48] Note that in agreement with numerical simulations, we only expect oscillations within the LD phase, since in both HD and MC, the current J is independent of the delay between loading and exit (see Supp. Mat.).
[49] M. Elowitz and S. Leibler, Nature **403**, 335 (2000).
[50] B. Ananthasubramaniam and H. Herzel, PLoS ONE **2014**, e104761 (2014).
[51] S. Oliveto, M. Mancino, N. Manfrini, and S. Biffo, World J Biol Chem **8**, 45?56 (2017).
[52] S. Kojima, E. Sher-Chen, and C. Green, Gene Dev. **26**, 2724 (2012).

Translational control of gene expression via interacting feedback loops. Supplemental Material.

Liang Wang
University of Dundee, Dundee, United Kingdom

M. Carmen Romano
University of Aberdeen, Aberdeen, United Kingdom

Fordyce A. Davidson*
University of Dundee, Dundee, United Kingdom
Corresponding Author

October 8, 2019

1 Protein dynamics and effective vacancy of the Start Site

Protein complex formation and loss respectively can be represented as follows:

$$\underbrace{\mathrm{P}+\mathrm{P}+\ldots\mathrm{P}}_{n} \underset{r_2}{\overset{r_1}{\rightleftharpoons}} n\mathrm{P}, \quad \mathrm{P}\xrightarrow{r}\emptyset$$

where P is protein, $n \geq 1$ proteins form a protein complex nP and r_1, r_2 and r are rate constants with dimension proportional to 1/time. Applying a quasi-steady-state approximation for the complex concentration:

$$\frac{d[n\mathrm{P}]}{dt} = r_1\left([\mathrm{P}]\right)^n - r_2[n\mathrm{P}] \approx 0 \quad \Longrightarrow \quad [n\mathrm{P}] = \frac{r_1}{r_2}\left([\mathrm{P}]\right)^n = \frac{r_1}{r_2}\left(\frac{N}{V}\right)^n, \tag{S1}$$

where $[\,\cdot\,]$ represents concentration, N denotes the number of molecules of P, and V denotes the cell volume. The protein complex is assumed to bind and unbind from some location at the start site (SS) at rates k_1, k_2, respectively:

$$n\mathrm{P} + \mathrm{SS} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} n\mathrm{P}\cdot\mathrm{SS}. \tag{S2}$$

Writing (S2) as a differential equation and applying the quasi-steady state approximation as above yields

$$\begin{aligned}\frac{d[n\mathrm{P}\cdot\mathrm{SS}]}{dt} &= k_1 \times [n\mathrm{P}] \times [\mathrm{SS}] - k_2 \times [n\mathrm{P}\cdot\mathrm{SS}] \approx 0 \\ i.e. \quad [n\mathrm{P}\cdot\mathrm{SS}] &= \frac{k_1}{k_2}[n\mathrm{P}][\mathrm{SS}].\end{aligned} \tag{S3}$$

This quasi-steady state approximation leads to an *effective vacancy* of the SS from the perspective of ribosome binding events. The probability that the SS is vacant is equivalent to

$$\frac{[\mathrm{SS}]}{[\mathrm{SS}] + [n\mathrm{P}\cdot\mathrm{SS}]} = \frac{[\mathrm{SS}]}{[\mathrm{SS}] + \frac{k_1}{k_2}[n\mathrm{P}][\mathrm{SS}]} = \frac{k_2}{k_2 + k_1[n\mathrm{P}]}, \tag{S4}$$

where we have used (S3). Using (S1) and (S4) this probability can be written as

$$f(N) := \frac{1}{1 + (N/\theta)^n}, \tag{S5}$$

where $\theta = V\sqrt[n]{r_2k_2/r_1k_1}$.

2 Parameters

We rescale the time by the ribosome hopping rate γ. Using values for *E.coli* as a typical example, we take the average hopping rate is 22 amino acids per second [2]. Therefore, we consider the unit used in our simulations to be $U = \frac{1}{22}$ s. It has been reported that around 80% of all mRNAs have a half-lives, T_h, between 3 and 8 min [1]. We take the average of these figures, namely $T_h = 5.5$ min. Now consider a general system variable X that is subject to exponential decay with rate d:

$$\frac{dX(t)}{dt} = -dX(t) \quad \Longrightarrow \quad X(t) = X_0 e^{-dt}, \qquad X_0 = X(0). \tag{S6}$$

Let T_m be mean life of X and T_h be half life. Knowing that T_m is $1/d$, we are able to solve T_h by:

$$X(T_h) = X_0 e^{-dT_h} = \frac{1}{2} X_0, \tag{S7}$$

and

$$T_h = (\ln 2)/d = T_m \ln 2. \tag{S8}$$

Therefore the mean life time of mRNA is taken as $T_m = 5.5/\ln 2$ min. In terms of the simulation time units this becomes

$$\frac{5.5}{\ln 2}\,\text{min} = \frac{5.5}{\ln 2} \times 60\,\text{s} = \frac{5.5}{\ln 2} \times 60 \times 22\,\text{U} \approx 10^4\,\text{U}. \tag{S9}$$

The average length of mRNA in prokaryotic cells is about $1000 - 1500$ nucleotides. Each codon consist of three nucleotides, so average length of prokaryotic mRNA is about $333 - 500$ codons. We use $L = 500$ in our simulations.

3 Phases and phase boundaries

By direct comparison with the standard TASEP it is proposed that in the models presented here, steady state values of J and hence N can be segregated into different phases defined as follows. In each case

$$\alpha_{eff} := \alpha f(N) + k\rho_L, \qquad \beta_{eff} := \beta + k(1 - \rho_1). \tag{S10}$$

[Note that for the competitive recycling case, the effective rates are given by $\alpha^c_{eff} := (\alpha + k\rho_L) f(N)$ and $\beta^c_{eff} := \beta + f(N)k(1 - \rho_1)$.]

Maximal Current Phase: α_{eff}, $\beta_{eff} \geqslant 1/2$

$$\rho_1 = 1 - \frac{1}{4\alpha_{eff}}, \qquad \rho_L = \frac{1}{4\beta_{eff}}, \qquad \rho = \frac{1}{2}, \qquad J = \frac{1}{4}. \tag{S11}$$

Low Density Phase: $\alpha_{eff} < \beta_{eff}$, $\alpha_{eff} < 1/2$

$$\rho_1 = \alpha_{eff}, \qquad \rho_L = \frac{\alpha_{eff}(1 - \alpha_{eff})}{\beta_{eff}}, \qquad \rho = \alpha_{eff}, \qquad J = \alpha_{eff}(1 - \alpha_{eff}). \tag{S12}$$

High Density Phase: $\alpha_{eff} > \beta_{eff}$, $\alpha_{eff} > 1/2$

$$\rho_1 = 1 - \frac{\beta_{eff}(1 - \beta_{eff})}{\alpha_{eff}}, \qquad \rho_L = 1 - \beta_{eff}, \qquad \rho = 1 - \beta_{eff}, \qquad J = \beta_{eff}(1 - \beta_{eff}). \tag{S13}$$

The corresponding steady state protein level is given by $N^* := J/r$.

Lemma 1 *The Maximal Current Phase is defined by*

$$\frac{\alpha}{2\alpha + k\left(1 + I^n\right)} \leqslant \begin{cases} \beta & \leqslant \dfrac{\alpha k}{1 + I^n - 2\alpha} \quad \textit{if} \quad \alpha < \dfrac{1}{2}\left(1 + I^n\right), \\ \beta & \textit{otherwise}, \end{cases} \tag{S14}$$

where $I := \frac{1}{4r\theta}$. *Within this region,*

$$\alpha_{eff} = \frac{\alpha}{1+I^n} + \frac{k}{2\left(\beta + \sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1+I^n\right)\right)}\right)} \quad \textit{and} \quad \beta_{eff} = \frac{1}{2}\left(\beta + \sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1+I^n\right)\right)}\right). \tag{S15}$$

Remark 1 *Note that the stacking of the left and right inequalities* (S14) *in case* $\alpha < \frac{1}{2}\left(1+I^n\right)$ *is necessary and therefore generates a lower bound for* α *and* β *marked by the intersection of the two functions of* α. *It is straight forward to compute this intersection:*

$$(\alpha, \beta) = \left(\frac{1-k}{2}\left(1+I^n\right), \frac{1-k}{2}\right). \tag{S16}$$

Lemma 2 *The Low Density Phase is defined by*

$$\alpha < \begin{cases} \beta\left(1 + \left(\frac{4I(\beta+k)(1-\beta)}{(1+k)^2}\right)^n\right) & \textit{if} \quad 2\beta + k < 1, \\ \frac{\beta}{2\beta+k}\left(1+I^n\right) & \textit{otherwise.} \end{cases} \tag{S17}$$

Within this region, there exists a unique, positive expression for α_{eff} *that yields unique, positive expressions for* β_{eff} *and* N^*.

Lemma 3 *The High Density Phase is defined by*

$$\alpha > \begin{cases} \beta\left(1 + \left(\frac{4I(\beta+k)(1-\beta)}{(1+k)^2}\right)^n\right) & \textit{if} \quad 2\beta + k < 1, \\ \frac{k\beta}{1-2\beta}\left(1+I^n\right) & \textit{otherwise,} \end{cases} \qquad \beta < \frac{1}{2}. \tag{S18}$$

Within this region, there exist either one or three positive solutions, β_{eff}, *that yield positive expressions for* α_{eff} *and* N^*.

4 Proofs

Proof of Lemma 1 Substituting the effective rates given in Eq. 2 main text into the definitions Eq 4 main text yields:

$$\beta_{eff} = \beta + k(1-\rho_1) = \beta + k\left(1 - \left(1 - \frac{1}{4\alpha_{eff}}\right)\right) = \beta + \frac{k}{4}\frac{1}{\alpha_{eff}}.$$

Therefore, we obtain a quadratic equation for β_{eff}:

$$4^{n+1}\alpha(\theta r)^n\beta_{eff}^2 - 4^{n+1}\alpha\beta(\theta r)^n\beta_{eff} - k\beta\left((4\theta r)^n + 1\right) = 0.$$

Subsequently,

$$\beta_{eff} = \frac{1}{2}\left(\beta \pm \sqrt{\beta^2 + \frac{k\beta\left((4\theta r)^n+1\right)}{4^n\alpha(\theta r)^n}}\right).$$

Since we require β_{eff} to be positive and $\sqrt{\beta^2 + \frac{k\beta\left((4\theta r)^n+1\right)}{4^n\alpha(\theta r)^n}} > \beta$, it follows that:

$$\beta_{eff} = \frac{1}{2}\left(\beta + \sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1 + \left(\frac{1}{4\theta r}\right)^n\right)\right)}\right). \tag{S19}$$

Then α_{eff} is easy to obtain:

$$\alpha_{eff} = \frac{\alpha}{1+\left(\frac{1}{4\theta r}\right)^n} + \frac{k}{2\left(\beta + \sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)\right)}\right)}. \tag{S20}$$

The maximal current phase is determined by the conditions α_{eff}, $\beta_{eff} \geqslant \frac{1}{2}$ i.e.:

$$\beta_{eff} = \frac{\beta + \sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)\right)}}{2} \geqslant \frac{1}{2},$$

so $\sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)\right)} \geqslant 1-\beta$. Since $1-\beta > 0$ for $\beta < 1$, we have

$$\beta \geqslant \frac{\alpha}{2\alpha + k\left(1+\left(\frac{1}{4\theta r}\right)^n\right)}, \tag{S21}$$

and

$$\alpha_{eff} = \frac{\alpha}{1+\left(\frac{1}{4\theta r}\right)^n} + \frac{k}{2\left(\beta + \sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)\right)}\right)} \geqslant \frac{1}{2},$$

namely,

$$\frac{k}{2\left(\beta + \sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)\right)}\right)} \geqslant \frac{1}{2} - \frac{\alpha}{1+\left(\frac{1}{4\theta r}\right)^n}. \tag{S22}$$

If $\frac{1}{2} - \frac{\alpha}{1+\left(\frac{1}{4\theta r}\right)^n} \leqslant 0$, i.e., $\alpha \geqslant \frac{1}{2}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)$, then (S22) is always true.

If $\alpha < \frac{1}{2}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)$, we have:

$$k \geqslant 2\left(\frac{1}{2} - \frac{\alpha}{1+\left(\frac{1}{4\theta r}\right)^n}\right)\left(\beta + \sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)\right)}\right),$$

subsequently,

$$k - \left(1 - \frac{2\alpha}{1+\left(\frac{1}{4\theta r}\right)^n}\right)\beta \geqslant \left(1 - \frac{2\alpha}{1+\left(\frac{1}{4\theta r}\right)^n}\right)\sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)\right)} > 0. \tag{S23}$$

From (S23), we have a weak condition

$$\beta \leqslant \frac{\alpha k}{1+\left(\frac{1}{4\theta r}\right)^n - 2\alpha}. \tag{S24}$$

The result follows by grouping the inequalities determined above and on writing the final expression in terms of the intensity factor.

□

Proof of Lemma 2 Substituting the effective rates given in Eq. 2 main text into the definitions Eq. 5 main text yields:

$$\alpha_{eff} = \frac{\alpha}{1+\left(\frac{\alpha_{eff}(1-\alpha_{eff})}{\theta r}\right)^n} + k\frac{\alpha_{eff}(1-\alpha_{eff})}{\beta + k(1-\alpha_{eff})}.$$

Again after some manipulation we obtain:

$$P(\alpha_{eff}) := \beta\alpha_{eff}^{n+1}(1-\alpha_{eff})^n + (\theta r)^n(\beta+\alpha k)\alpha_{eff} - \alpha(\theta r)^n(\beta+k) = 0. \tag{S25}$$

If there exist positive solutions, α_{eff}, to (S25), then from (S12), $\beta_{eff} = \beta + k\,(1-\alpha_{eff})$ and $N^* = \alpha_{eff}(1-\alpha_{eff})/r$. Next we consider the conditions for the existence of α_{eff} and thus define the phase boundary for the LD phase. LD phase is by defined by $\alpha_{eff} < \beta_{eff}$, $\alpha_{eff} < \frac{1}{2}$. From $\alpha_{eff} < \beta_{eff}$, we have:

$$\alpha_{eff} < \beta_{eff} = \beta + k(1-\alpha_{eff}), \quad \Longrightarrow \quad (1+k)\alpha_{eff} < \beta + k,$$

therefore

$$\alpha_{eff} < \frac{\beta+k}{1+k}. \tag{S26}$$

On the other hand, we require $\alpha_{eff} < \frac{1}{2}$. Therefore $\alpha_{eff} < \min\left(\frac{1}{2}, \frac{\beta+k}{1+k}\right)$.

If $\frac{\beta+k}{1+k} \geqslant \frac{1}{2}$, i.e., $2\beta + 2k \geqslant 1+k$, $2\beta + k \geqslant 1$, then $\alpha_{eff} < \frac{1}{2}$. The derivative of P with respect to α_{eff} is

$$P' = \beta\alpha_{eff}^n(1-\alpha_{eff})^{n-1}\left(1-\alpha_{eff}+n(1-2\alpha_{eff})\right) + (\theta r)^n(\beta+\alpha k).$$

Since both $1-\alpha_{eff}$ and $1-2\alpha_{eff}$ are positive in $\left(0, \frac{1}{2}\right)$, so $P' > 0$ in $\left(0, \frac{1}{2}\right)$. Moreover, $P(0) = -\alpha(\theta r)^n(\beta+k) < 0$. Therefore, a real, positive solution of (S25) exists if and only if $f\left(\frac{1}{2}\right) > 0$ (in this case, the solution is unique). Now,

$$P\left(\frac{1}{2}\right) = \frac{(\theta r)^n}{2}\left(\beta\left(1+\left(\frac{1}{4\theta r}\right)^n\right) - \alpha(2\beta+k)\right),$$

so $P\left(\frac{1}{2}\right) > 0$ requires $\beta\left(1+\left(\frac{1}{4\theta r}\right)^n\right) - \alpha(2\beta+k) > 0$, namely,

$$\alpha < \frac{\beta}{2\beta+k}\left(1+\left(\frac{1}{4\theta r}\right)^n\right). \tag{S27}$$

If $\frac{\beta+k}{1+k} < \frac{1}{2}$, i.e., $2\beta + k < 1$, then $\alpha_{eff} < \frac{\beta+k}{1+k}$. Since P is increasing in $\left(0, \frac{1}{2}\right)$ and $P(0) < 0$, so a solution of (S25) exists if and only if $P\left(\frac{\beta+k}{1+k}\right) > 0$ (and again the solution is unique).

$$P\left(\frac{\beta+k}{1+k}\right) = \frac{(\beta+k)(\theta r)^n}{1+k}\left(\beta\left(\frac{(\beta+k)(1-\beta)}{\theta r(1+k)^2}\right)^n + \beta - \alpha\right) > 0,$$

provided

$$\alpha < \beta\left(1+\left(\frac{(\beta+k)(1-\beta)}{\theta r(1+k)^2}\right)^n\right). \tag{S28}$$

The proof is complete by writing the inequalities in terms of the Intensity Factor.

□

Proof of Lemma 3 Substituting the effective rates given in Eq. 2 main text into the definitions Eq. 6 main text yields:

$$\beta_{eff} = \beta + k\frac{\beta_{eff}(1-\beta_{eff})}{\dfrac{\alpha}{1+\left(\dfrac{N}{\theta}\right)^n} + k\rho_L} = \beta + \frac{k\beta_{eff}(1-\beta_{eff})}{\dfrac{\alpha}{1+\left(\dfrac{\beta_{eff}(1-\beta_{eff})}{\theta r}\right)^n} + k(1-\beta_{eff})}.$$

Further algebra yields the equation

$$Q(\beta_{eff}) := k\beta\beta_{eff}^n\left(1-\beta_{eff}\right)^{n+1} - (k\beta+\alpha)(\theta r)^n\beta_{eff} + \beta(\theta r)^n(\alpha+k) = 0. \tag{S29}$$

With β_{eff} a solution of (S29), then from (S13), $\rho = 1-\beta_{eff}$, $J = \beta_{eff}(1-\beta_{eff})$ and $N^* = \beta_{eff}(1-\beta_{eff})/r$. Hence, the effective initiation rate is given by

$$\alpha_{eff} = \frac{\alpha}{1+\left(\dfrac{\beta_{eff}(1-\beta_{eff})}{\theta r}\right)^n} + k\left(1-\beta_{eff}\right).$$

One of the HD conditions is $\beta_{eff} < \dfrac{1}{2}$, i.e. $\dfrac{1}{2} > \beta_{eff} = \beta + k(1-\rho_1) > \beta$, namely, $\beta < \dfrac{1}{2}$. Moreover, from the second condition for HD $\alpha_{eff} > \beta_{eff}$, we have

$$\beta_{eff} = \beta + k\frac{\beta_{eff}(1-\beta_{eff})}{\alpha_{eff}} < \beta + k(1-\beta_{eff}) \quad \Longrightarrow \quad (1+k)\beta_{eff} < \beta + k.$$

Therefore, we require $\beta_{eff} < \min\left\{\dfrac{1}{2}, \dfrac{\beta+k}{1+k}\right\}$.

If $\dfrac{\beta+k}{1+k} \geqslant \dfrac{1}{2}$, i.e., $2\beta + k \geqslant 1$ then we need to check the existence of solutions to (S29) in $\left(0, \dfrac{1}{2}\right)$. To this end, note that

$$\begin{aligned} Q' &= k\beta\beta_{eff}^{n-1}\left(1-\beta_{eff}\right)^n\left(n-(2n+1)\beta_{eff}\right) - (k\beta+\alpha)(\theta r)^n, \\ Q'' &= nk\beta\beta_{eff}^{n-2}\left(1-\beta_{eff}\right)^{n-1}\left(2(2n+1)\beta_{eff}^2 - 4n\beta_{eff} + n - 1\right). \end{aligned}$$

Thus, $Q'' > 0$ in $\left(0, \beta_{eff}^-\right)$ and $Q'' < 0$ in $\left[\beta_{eff}^-, \dfrac{1}{2}\right)$ with the maximum of Q' obtained at β_{eff}^-. Moreover,

$$Q'(0) = -(k\beta+\alpha)(\theta r)^n < 0 \quad \text{and} \quad Q'\left(\frac{1}{2}\right) = -k\beta\frac{1}{4^n} - (k\beta+\alpha)(\theta r)^n < 0.$$

If $Q'\left(\beta_{eff}^-\right) < 0$, then $Q' < 0$ (Q is decreasing) in $\left(0, \dfrac{1}{2}\right)$. Besides, $Q(0) = \beta(\theta r)^n(\alpha+k) > 0$. Thus, the solution of (S29) $Q = 0$ exists if and only if $Q\left(\dfrac{1}{2}\right) < 0$, and the solution is unique.

If $Q'\left(\beta_{eff}^-\right) > 0$, then there are two solutions to $Q' = 0$. Let β_{eff}^1 and β_{eff}^2 denote the solutions and $0 < \beta_{eff}^1 < \beta_{eff}^2 < \dfrac{1}{2}$. Thus, $Q' < 0$ (Q is decreasing) in $\left(0, \beta_{eff}^1\right) \cup \left(\beta_{eff}^2, \dfrac{1}{2}\right)$, $Q' > 0$ (Q is increasing) in $\left(\beta_{eff}^1, \beta_{eff}^2\right)$. Moreover, $Q(0) = \beta(\theta r)^n(\alpha+k) > 0$, and $Q\left(\beta_{eff}^1\right)$ is the local minimum. We have following cases.

(I) If $Q\left(\beta_{eff}^1\right) < 0$ and $Q\left(\beta_{eff}^2\right) < 0$, then $Q < 0$ in $\left(\beta_{eff}^1, \dfrac{1}{2}\right)$, there exists a unique solution to (S29) in $\left(0, \beta_{eff}^1\right)$. In this case, $Q\left(\dfrac{1}{2}\right) < 0$.

(III) If $Q\left(\beta_{eff}^1\right) < 0$ and $Q\left(\beta_{eff}^2\right) > 0$, then (S29) has three solutions if $Q\left(\frac{1}{2}\right) < 0$ or two solutions if $Q\left(\frac{1}{2}\right) > 0$. If $Q\left(\frac{1}{2}\right) > 0$, we have:

$$Q\left(\frac{1}{2}\right) = \frac{(\theta r)^n}{2}\left(k\beta\left(\frac{1}{4\theta r}\right)^n - \alpha + 2\beta\alpha + \beta k\right) > 0,$$

which gives:

$$\beta > \frac{\alpha}{2\alpha + k\left(1 + \left(\frac{1}{4\theta r}\right)^n\right)}. \tag{S30}$$

However, from (S14), (S30) sets the system in the MC phase and hence is not a feasible condition. Therefore, in this case, $Q\left(\frac{1}{2}\right) < 0$ and there are three solutions to (S29).

Overall, for $2\beta + k \geqslant 1$, the existence of the solution of β_{eff} requires $Q\left(\frac{1}{2}\right) < 0$:

$$Q\left(\frac{1}{2}\right) = \frac{(\theta r)^n}{2}\left(k\beta\left(\frac{1}{4\theta r}\right)^n - \alpha + 2\beta\alpha + \beta k\right) < 0,$$

which gives:

$$\alpha > \frac{k\beta\left(1 + \left(\frac{1}{4\theta r}\right)^n\right)}{1 - 2\beta}. \tag{S31}$$

If $\frac{\beta + k}{1 + k} < \frac{1}{2}$, i.e., $2\beta + k < 1$, then we require $\beta_{eff} < \frac{\beta + k}{1 + k}$. Similar to the case for $2\beta + k \geqslant 1$, the existence of the solution of β_{eff} requires $Q\left(\frac{\beta + k}{1 + k}\right) < 0$ and one or three solutions can be obtained. Note,

$$Q\left(\frac{\beta + k}{1 + k}\right) = \frac{(\theta r)^n k(1 - \beta)}{1 + k}\left(\beta\left(\frac{(\beta + k)(1 - \beta)}{\theta r(1 + k)^2}\right)^n + \beta - \alpha\right) < 0,$$

provided

$$\alpha > \beta\left(1 + \left(\frac{(\beta + k)(1 - \beta)}{\theta r(1 + k)^2}\right)^n\right). \tag{S32}$$

The result follows directly as above. □

5 Simulation

In every Monte-Carlo step (MCS), $L + 2$ random numbers were chosen, corresponding to all possible reactions. Protein protein degradation was modelled using n_r (the current number of proteins) attempts, each with corresponding degradation rate, r. The initiation and termination probabilities were governed by α_{eff} and β_{eff}. The processes of complex formation and subsequent interaction with the Start Site were implemented using the quasi-steady state approximations detailed above.

Time average was calculated from a simulation consisting of 10^8 Monte-Carlo steps (MCS), where the first 10^6 MCS were disregarded to ensure steady-state. Then, the integration time was divided in windows of 10^4 MCS, so that the standard deviation could be calculated for each value of α.

6 Dynamics in the Modified TASEP

6.1 Onset of oscillations is characterised by a Hopf Bifurcation

The mean time for a particle to transit the lattice can be estimated as

$$T = \frac{L}{1 - \langle \rho \rangle} = \frac{L}{1 - \langle \alpha_{eff} \rangle} \tag{S33}$$

in the LD regime. After substitution and some rearranging, the effective ribosome binding rate in LD is

$$\alpha_{eff}(N) = \frac{\alpha(\beta + k)}{\alpha k + \beta \left(1 + (N/\theta)^n\right)}. \tag{S34}$$

Appealing to the mean field approximation and setting $\langle N \rangle = N^*$, the mean transit time is then defined as

$$\langle T \rangle = \frac{L}{1 - \frac{\alpha(\beta+k)}{\alpha k + \beta\left(1 + \left(\frac{\langle N \rangle}{\theta}\right)^n\right)}}.$$

The delay differential equation for protein copy number in the LD regime is:

$$\frac{dN(t)}{dt} = \frac{\alpha(\beta + k)}{\alpha k + \beta \left(1 + \left(\frac{N(t-T)}{\theta}\right)^n\right)} \left(1 - \frac{\alpha(\beta + k)}{\alpha k + \beta \left(1 + \left(\frac{N(t-T)}{\theta}\right)^n\right)}\right) - rN(t), \tag{S35}$$

where we have written T for $\langle T \rangle$ for ease of exposition here and below. Steady states, N^*, of (S35) are exactly the solutions generated by substituting solutions of (S25) into the definition $N^* = \alpha_{eff}(1 - \alpha_{eff})/r$.

Substituting $N(t) = N^* + \delta(t)$ with $|\delta| \ll 1$into (S35), using Taylor expansion and dropping the higher order term yields

$$\frac{d\delta(t)}{dt} = -B(N^*)\delta(t - T) - r\delta(t), \tag{S36}$$

where

$$B(N^*) := \left(1 - \frac{2\alpha(\beta + k)}{\alpha k + \beta \left(1 + \left(\frac{N}{\theta}\right)^n\right)}\right) \frac{\alpha\beta(\beta + k)n \left(\frac{N}{\theta}\right)^n}{N \left(\alpha k + \beta \left(1 + \left(\frac{N}{\theta}\right)^n\right)\right)^2} > 0 \tag{S37}$$

as $\alpha_{eff} < 1/2$ in the LD phase by definition. Setting $\delta(t) = Ae^{\lambda t}$ and on substitution into (S36), we get (hereon setting $B(N^*) = B$ for ease of notation)

$$\lambda + r = -Be^{-\lambda T}. \tag{S38}$$

Setting $\lambda = \mu + i\omega$, standard arguments yield a critical wave number at which stability is lost, namely, $\omega^2 = B^2 - r^2$ and hence a necessary condition is $B > r$. Some algebra eliminates ω to defines the Hopf locus in terms of system parameters thus:

$$B \cos\left(\sqrt{B^2 - r^2}\, T\right) + r = 0. \tag{S39}$$

Hence, the existence of periodic solutions to (S35) is guaranteed, at least for values of (α, β) sufficiently close to the locus. The Hopf locus in the $\alpha - \beta$ plane is shown in Figure S3 along with the effects of varying I and k (Figs. S3 A and B). Finally, note that the necessary condition $B - r > 0$, yields:

$$\frac{\alpha\beta(\beta + k)}{\left(\alpha k + \beta \left(1 + \left(\frac{N^*}{\theta}\right)^n\right)\right)^3} g\left(\left(\frac{N^*}{\theta}\right)^n\right) > 0,$$

where

$$g\left(\left(\frac{N^*}{\theta}\right)^n\right) = \beta(n-1)\left(\frac{N^*}{\theta}\right)^{2n} - (\beta+\alpha k)(1-\alpha) + (\beta(\alpha+n-2\alpha n-2)-(n+1)\alpha k)\left(\frac{N^*}{\theta}\right)^n. \tag{S40}$$

Hence, $B-r>0 \implies g\left(\left(\frac{N^*}{\theta}\right)^n\right)>0$. Notice that if $n=1$, $g\left(\left(\frac{N^*}{\theta}\right)^n\right) = (-(\alpha+1)\beta-2\alpha k)\frac{N^*}{\theta} - (\beta+\alpha k)(1-\alpha)<0$. Hence the Hopf bifurcation cannot occur in this case. Therefore, co-operativity is a necessary condition for Hopf bifurcation to occur.

For $n>1$, the discriminant of $g\left(\left(\frac{N^*}{\theta}\right)^n\right)$ is given by:

$$\Delta = (\beta(\alpha+n-2\alpha n-2)-(n+1)\alpha k)^2 + 4\beta(n-1)(\beta+\alpha k)(1-\alpha), \tag{S41}$$

and $\Delta>0$ certainly when $\alpha<1$. Therefore, in this case $g\left(\left(\frac{N^*}{\theta}\right)^n\right)=0$ has two distinct real roots:

$$\left(\frac{N^*}{\theta}\right)^n_{\pm} = \frac{-(\beta(\alpha+n-2\alpha n-2)-(n+1)\alpha k)\pm\sqrt{\Delta}}{2\beta(n-1)},$$

where Δ is given in (S41) and $\left(\frac{N^*}{\theta}\right)^n_-$ is negative. Hence, it is easy to conclude that $g\left(\left(\frac{N^*}{\theta}\right)^n\right)>0$ requires:

$$\left(\frac{N^*}{\theta}\right)^n > \left(\frac{N^*}{\theta}\right)^n_+ = \frac{-(\beta(\alpha+n-2\alpha n-2)-(n+1)\alpha k)+\sqrt{\Delta}}{2\beta(n-1)}. \tag{S42}$$

The inequality (S42) and setting the right hand side of (S35) to zero, together imply

$$\theta r < \frac{\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N^*}{\theta}\right)^n_+\right)}\left(1-\frac{\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N^*}{\theta}\right)^n_+\right)}\right)\frac{1}{\sqrt[n]{\left(\frac{N^*}{\theta}\right)^n_+}}. \tag{S43}$$

Rewriting this in terms of the Intensity Factor I yields the condition $I>F(\mathbf{\Lambda})$ where $F(\mathbf{\Lambda})$ is a multiple of the reciprocal of the right hand side of (S43).

6.2 Numerical simulations of the characteristic DDE

The DDE (S35) was solved numerically using the code `Matlab dde`. Typical behaviour for a range of values of α is shown in the Fig. S5 below.

6.3 Periodic oscillations are restricted to the LD Phase

In the MC phase, $J\equiv 1/4$ and α_{eff} is given by (S20) and thus the loading rate is independent of time. Therefore $N\equiv 1/4r$ is the unique asymptotic solution related to any non-zero initial data. Similarly, in the HD phase, $J=\beta_{eff}(1-\beta_{eff})$, with β_{eff} given by the solutions of (S29). In this case, $\beta_{eff}=\beta_{eff}(N(t))$ and thus $J=J(N(t))$ and the characteristic equation $dN/dt=J-rN$ is a scalar, ordinary differential equation with constant coefficients. Standard theory precludes periodic solutions. The behaviour of the characteristic equations therefore supports the numerical simulations to suggest that (almost) periodic solutions are only realised in the LD phase.

We defined an autocorrelation measure $ACF(\tau) := E[(N(t)-\mu)(N(t+\tau)-\mu)]/\sigma^2$, where μ is the average and σ the standard deviation of the time series $N(t)$ of the number of proteins. $N(t)$ means the time series at time t, and $N(t+\tau)$ the time series at time $t+\tau$. Fig. S5 plots the first maximum of $ACF(\tau)$ for $\tau>0$. For a maximum to be "well-defined", the autocorrelation function must cross the horizontal axis twice. If there is no well-defined maximum, we set the measure equal to 0. Hence, large values (red) in the heatmap indicate pronounced oscillations in the number of proteins.

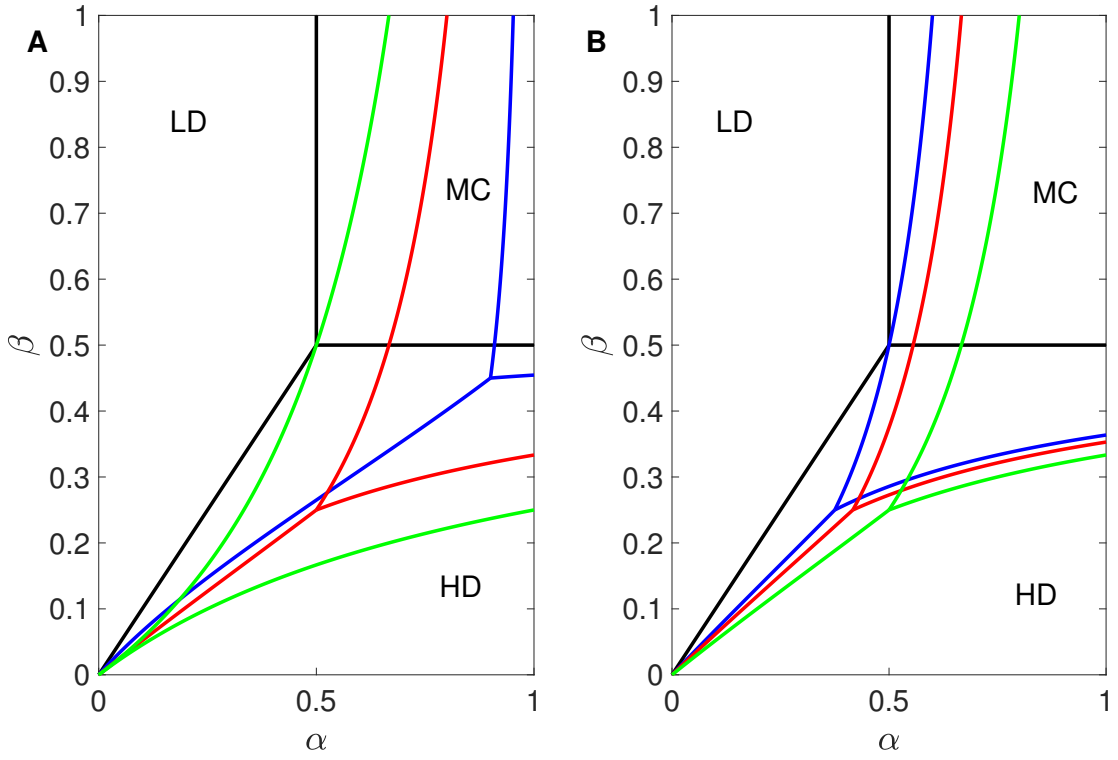

Figure S1: Phase diagrams for the modified TASEP. (A) Varying the recycling rate with fixed feedback intensity: $k = 0.1$ (blue), $k = 0.5$ (red), $k = 1$ (green) with $I = 1$. (B) (A) Varying the feedback intensity with fixed recycling rate: $I = 1$ (blue), $I = 2/3$ (red), $I = 0.5$ (green) with $k = 0.5$. The phase boundaries for the standard TASEP ($k = I = 0$) are shown in black. $L = 500$ sites, $r = 0.005$ in all cases

References

[1] JA Bernstein, AB Khodursky, PH Lin, S Lin-Chao, and SN Cohen. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proceeding of the National Academy of Sciences of the United States of America*, 99(15):9697–9702, 2002.

[2] ST Liang, YC Xu, P Dennis, and H Bremer. mRNA composition and control of bacterial gene expression. *Journal of Bacteriology*, 182(11):3037–3044, 2000.

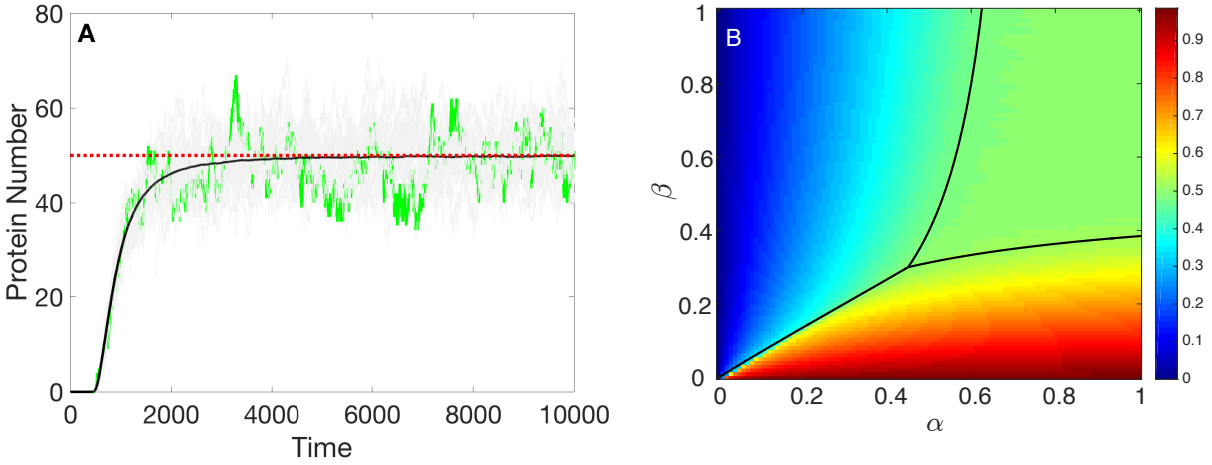

Figure S2: Typical phase diagram and associated simulation. (A) Monte-Carlo simulations as detailed in the text showing the temporal evolution of protein copy number ($\alpha = 0.8$ and $\beta = 0.5$ - MC phase). Grey lines are 20 typical trajectories with green line one example. Black line is the mean of 5000 simulations. Red dotted line is the steady state value predicted by the theory ($N^* = 50$). (B) Monte-Carlo simulations showing the average density of ribosomes on the mRNA depending on α and β. The black lines show the analytical mean-field phase boundaries derived above. The colour map indicates the average density, ρ. 5000 simulations were used for each representative (α, β)- pair. $L = 500$ sites, $n = 1$, $r = 0.005$, $I = 0.5$, $k = 0.4$

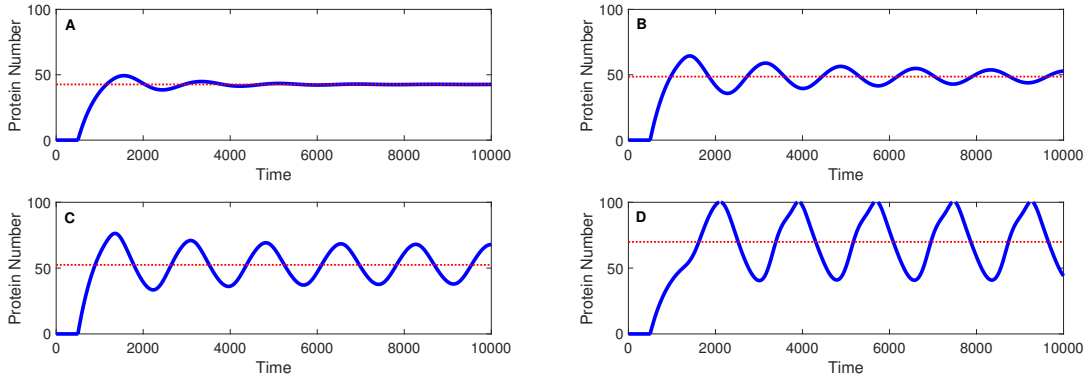

Figure S3: Onset of periodic oscillations in DDE simulation. Solutions of DDE (S35) showing onset of periodic solutions as α is increased through the Hopf locus shown in Figure S4. Blue lines show the numerical solution of (S35). The red dotted line shows the steady state value N^* as predicted by the steady state theory. (A) $\alpha = 0.1$, (B) $\alpha = 0.15$, (C) $\alpha = 0.2$, (D) $\alpha = 0.8$. Otherwise $L = 500$ sites, $\theta = 50$, $r = 0.002$, $k = 0.2$, $n = 5$, $\beta = 0.5$.

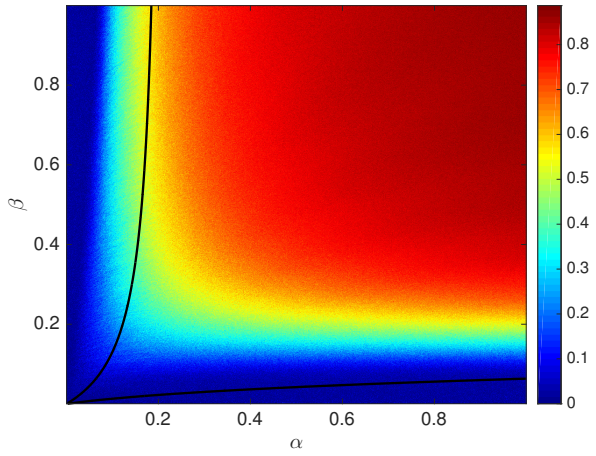

Figure S4: Periodicity in the Monte-Carlo simulations quantified by the autocorrelation function (heat map) compared to the Hopf bifurcation locus in the $\alpha - \beta$ plane analytically calculated using (S39) (upper black curve) and the LD/HD boundary from the mean-field approach (lower black curve). Heatmap shows an autocorrelation-based measure to numerically detect oscillations in the time series of the number of proteins. $I = 10/4$, $L = 500$ sites, $r = 0.002$, $k = 0.2$, $n = 5$.

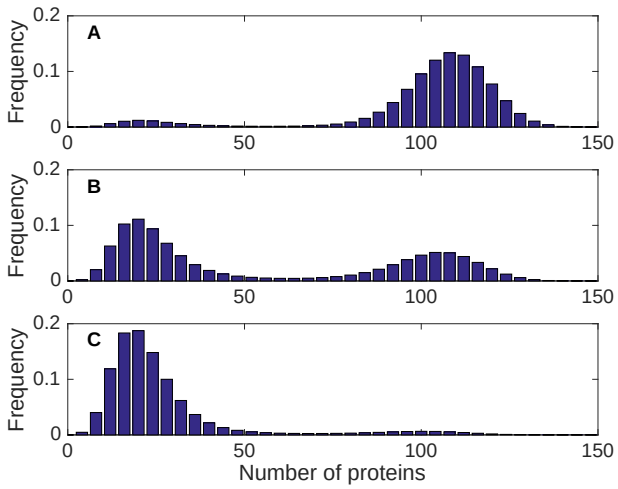

Figure S5: Histograms of the time series of the number of proteins in the bistable state for $L = 500$ sites, $\theta = 21$, $d = 0.002$, $k = 0.8$, $\beta = 0.015$ and $n = 2$ and three different values of α: (A) $\alpha = 0.75$, (B) $\alpha = 0.77$ and (C) $\alpha = 0.79$.

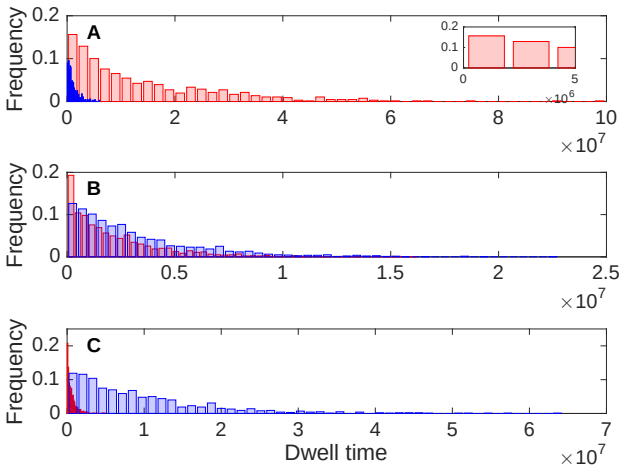

Figure S6: Histograms of the dwell times in each of the bistable states, calculated from the time series of the number of proteins in the bistable region, for $L = 500$ sites, $\theta = 21$, $d = 0.002$, $k = 0.8$, $\beta = 0.015$ and $n = 2$ and three different values of α: (A) $\alpha = 0.75$, (B) $\alpha = 0.77$ and (C) $\alpha = 0.79$. The red bars indicate the histogram of the dwell times in the upper stable state, and the blue bars the histogram of the dwell times in the lower stable state. The inset in panel (A) shows a zoom of the histogram for the upper stable state for low dwell times, to demonstrate that the smallest dwell times are larger than 0.

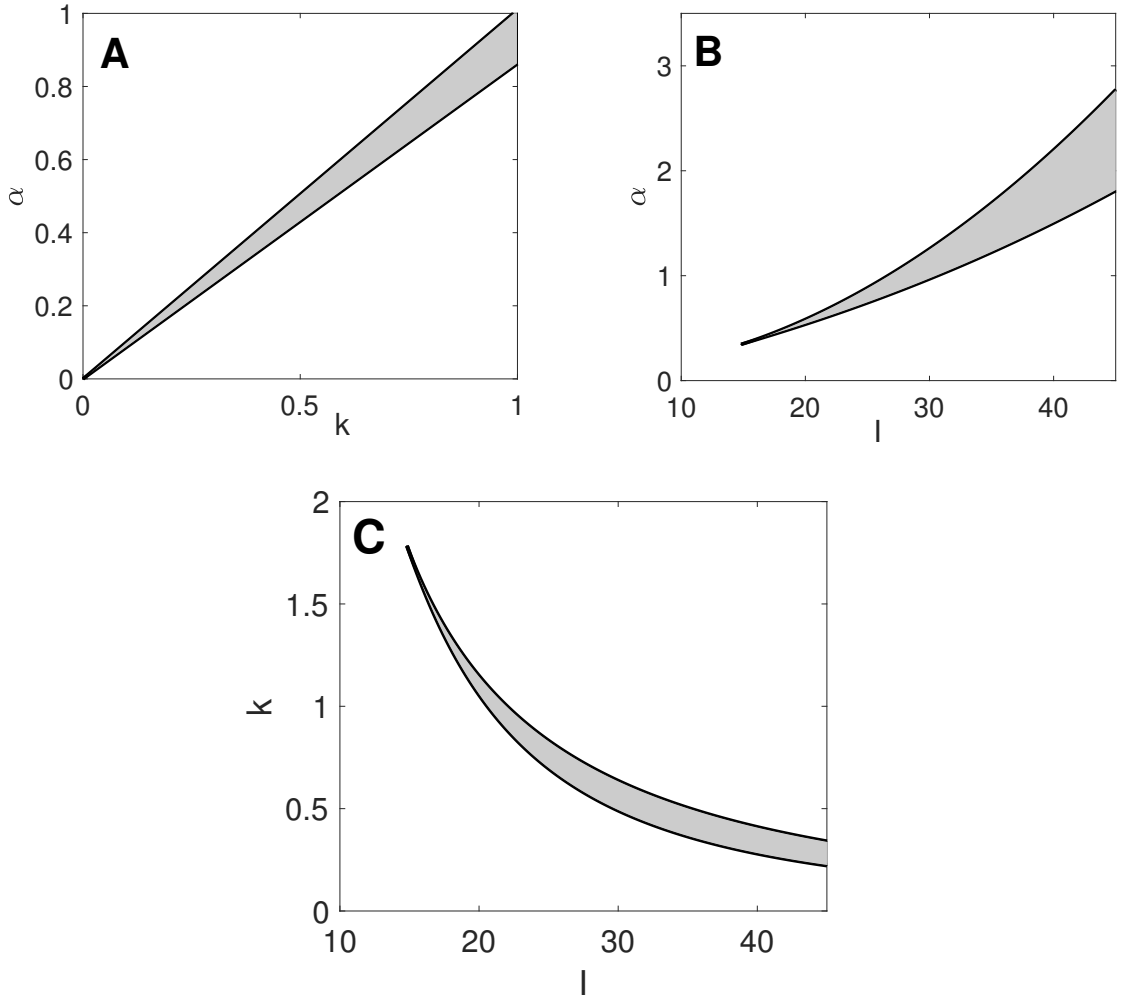

Figure S7: Regions of bistability as predicted by the steady state theory shown as functions of feedback (I), recycling (k) and loading rates (α). Bistable regions are shaded grey. (A) α-I plane ($k = 0.8$) (B) α-k plane ($I = 24$) (C) I-k plane ($\alpha = 0.77$). Otherwise $L = 500$, $n = 2$, $\beta = 0.015$.